Attachment 1

Ancillary Data Pre-processing

1. Introduction

Much processing of the instrument data acquired from EOS platfor require the input of ancillary data, both from within the environment (i.e. products from other EOS instruments), and from sources (e.g. NMC, NESDIS, etc.). It is not expected that ancill external source will be in a format suitable for immediate use by processing algorithms. The current ECS architecture has not pl constraints on where this pre-processing is performed (e.g. upon during data staging prior to processing, or even using specialized PDPS). However, the current scope of the ECS, in particular the SD data access toolkit tools, only provides science algorithms with from disc in an already pre-processed form, and does not and canno the issue of how the data are made available to the PDPS.

2. Scope of the Work

Many of the NMC data sets are gridded on a global basis and so are relatively simply after an initial reformatting. However, it is many NESDIS products have will need further preparation in terms o

- subsetting of specific parameters to reduce volume;
- re-presentation of complex swath structures

An extensive survey of the scientific user community was carri January '95. It was clear from this survey that there was unanim for common ancillary data sets (either specific data sets, or formats) being handled by ECS. We are now in a position to ident ancillary data types that are required by multiple users in an Release A time frame.

1- reformatting of NMC GRIB formatted products ,

2- reformatting of the NESDIS Snow/Ice Product operational pro

3- reformatting of TOMS products.

We know enough about tasks 1 & 2 to be confident that we can imp them as soon as a design is available, and we need to start work o clarify the requirements.

There are subsidiary data sets (in particular NMC BUFR formatted other NESDIS products) for which user commonality cannot be confir

---

[1]There is a clear requirement for this dataset. Currently the Level of the p

the scientific user community at this stage in their algorithm d
and consequently they are not scoped within this proposal.

3. Work Breakdown

It is proposed to divide the work into the following five activit

- Planning and Investigation
- Design and Implementation Phase I
- Design and Implementation Phase II
- Future Scoping
- Integration and Testing

In addition, the following lifecycle support activities are also re

- Regression Testing
- Sustaining Engineering
- Configuration Management
- Quality Assurance
- Technical Management

The schedule for each activity is given in Table 1. A key dri
proposed schedule is the need to support science algorithm bet
testing at the IR-1 DAACs in early '96.

3.1 Planning & Investigation

Although at this stage most of the data inputs are known, there a
of issues which need to be clarified before a design can be starte
investigations have to be initiated on all of the following issues

- Complete specification of the identified products;
- Specification of the output data types (proposed to be the evolving EOS-HDF standard);
- The extent of pre-processing & metadata generation;
- Defining Phase II capability;

It should be emphasized here that the scope of the proposed wo
provide only that level of metadata necessary to enable access b
toolkit.

We have scheduled this initial work to last for approximately thr
During this time the Ancillary Data Pre-processor Lead Engineer

closely with system engineering support to SDPS and the EOS Scie
Software Teams. The on-going engineering support needs to enhance
the current architecture to incorporate ancillary data preprocessi
EOS Science Software Teams will be needed to better identify
processing specifics. There are three possible places where ancill
processing can be accommodated within the current ECS architecture:

- integrated processes within the ingest client,
- type services within the data service, or
- specialized PGEs that work within PDPS

SDPS segment engineering will continue to analyze the trade-offs
these various solutions, in parallel with the initial investigati
new work.

3.2 Design and Implementation

We already have sufficient knowledge to perform 80% of
implementation. However, we cannot afford to delay starting work
pre-processing design & implementation until investigation is
This should not be seen as a potential risk, as the critical desig
to the ECS architecture, generic data types & formats (i.e. EOS-HDF
knowledge, etc. are all essentially internal to the ECS project,
resolved despite uncertainties in the final input data sets. In
need to specify interfaces, and provide functional software for
algorithm testing is now of critical concern. It is, therefore,
two phase development is under taken for the currently identified
capabilities.

Phase I will address the following software issues:-

- system architecture, ensuring a solution sufficiently flexi
  with both the Phase I and Phase II data set requirements;
- the capability to pre-process the NMC GRIB formatted pro
  and;
- interface them to the SDP toolkit ancillary data handling;
- integration of preliminary ancillary data QA and met
  functions.

Phase II will complete the currently identified Release A
requirements. Specifically it will provide:-

- development & integration of completed ancillary data QA
  metadata functions;
- the capability to pre-process & interface NESDIS Snow/Ice
  & TOMS products.

This two step approach enables us to start the major design proces
known data sets, and gives the other ECS teams longer to define

product inputs without placing the development schedule at risk. I
it also provides for an early availability of ancillary data to the
integration and evaluation.

It is estimated that 4,750 lines of code will need to be developed
software sizing estimate).

3.3 Future - Scoping

By the time initial work is complete, we shall be in a much bett
scope the remainder of the Release A, and the Release B developme
The work to be scoped in this package include:-

- introduction of new data types & formats;
- development of specialized pre-processing algorithms;
- further development of ancillary data QA and metad functionality;
- handling of low priority data sets.

3.4 Lifecycle Support Activities

Integration and test includes the effort required to integrate this
other Release A components, write test plans and procedures, dev
drivers and test data, conduct formal testing at the EDF and accept
at the DAACs, and write test reports.

Regression testing includes the effort required to retest this sc
Releases B, C, and D integration and test periods.

Sustaining engineering includes the effort required to perform
maintenance and nominal enhancements to this software, foll
delivery, for the remainder of the contract.

Configuration management includes the effort required to provide
configuration management services from the time the software en
Release A integration and test process through the end of the cont

Quality assurance includes the effort required to provide qualit
functions during design, development, integration and test, and
engineering for this software.

Management includes the effort required to provide project and
management of the design, development, integration and test, and s
engineering for this software.

These activities were estimated based on the size of the software effort from cost models used for the proposal and Change Order #1.

4.0 Effort Level

The proposed level of effort is given in Table 1.

Table 1: Proposed Effort Levels

Task	Start	End	Effort /man-months
Planning & Investigation	Mar‘95	Jun‘95	3.50
Phase I& II	Mar ‘95	Apr. ‘96	19.00
Future Scoping	Apr. ‘96	May .‘96	1.00
I&T	Nov.’ 95	May. ‘96	17.29
Regression Test	Jan. ‘97	Jun. ‘01	2.66
Sustaining Eng	Jan. ‘97	Oct. ‘02	9.12
CM	Nov. ‘95	Oct. ‘02	5.54
QA	Mar. ‘95	Oct. ‘02	0.41
Management	Mar. ‘95	Oct. ‘02	2.93

This gives a total of 61.45 labor-months effort between March October 2002.